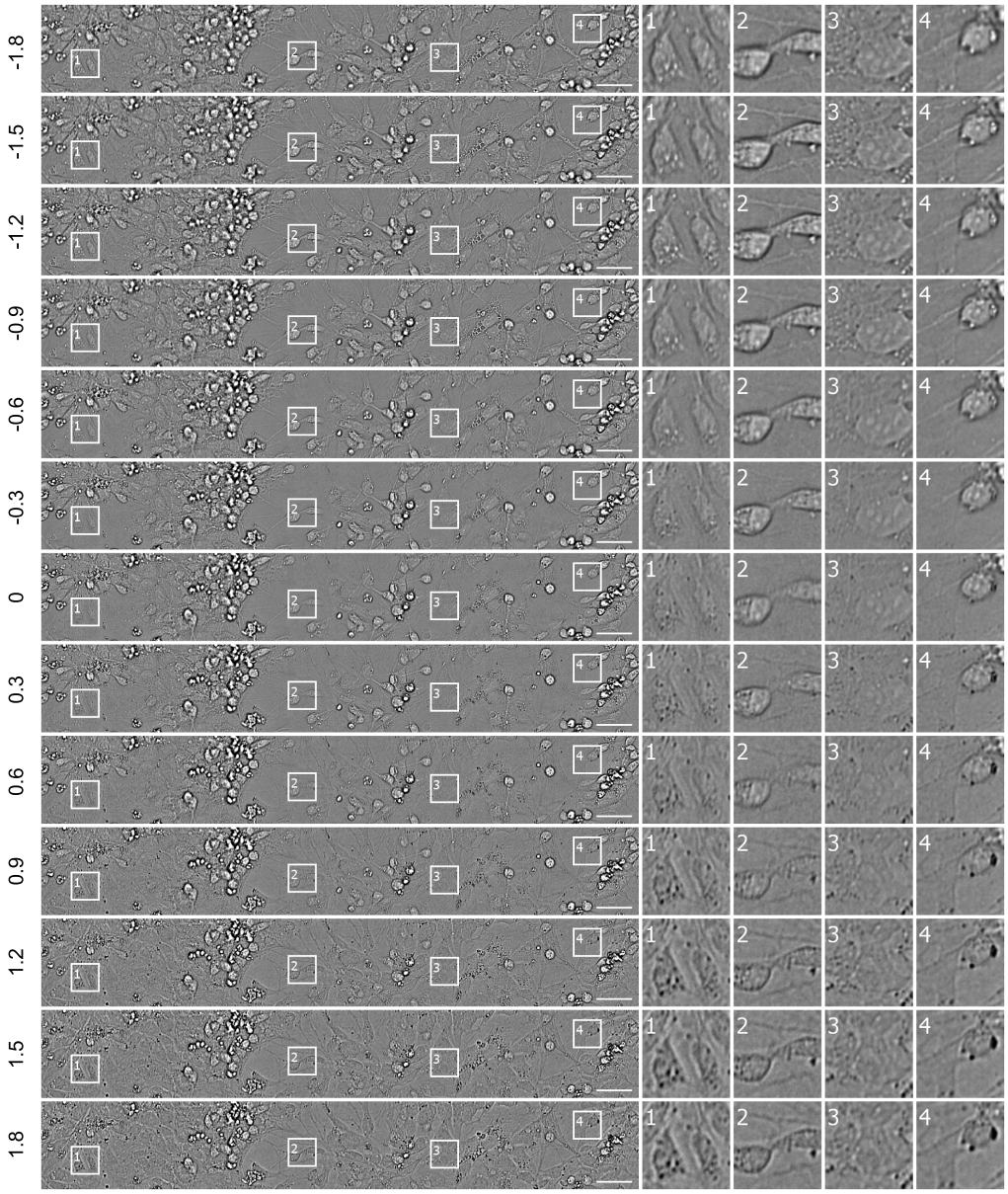
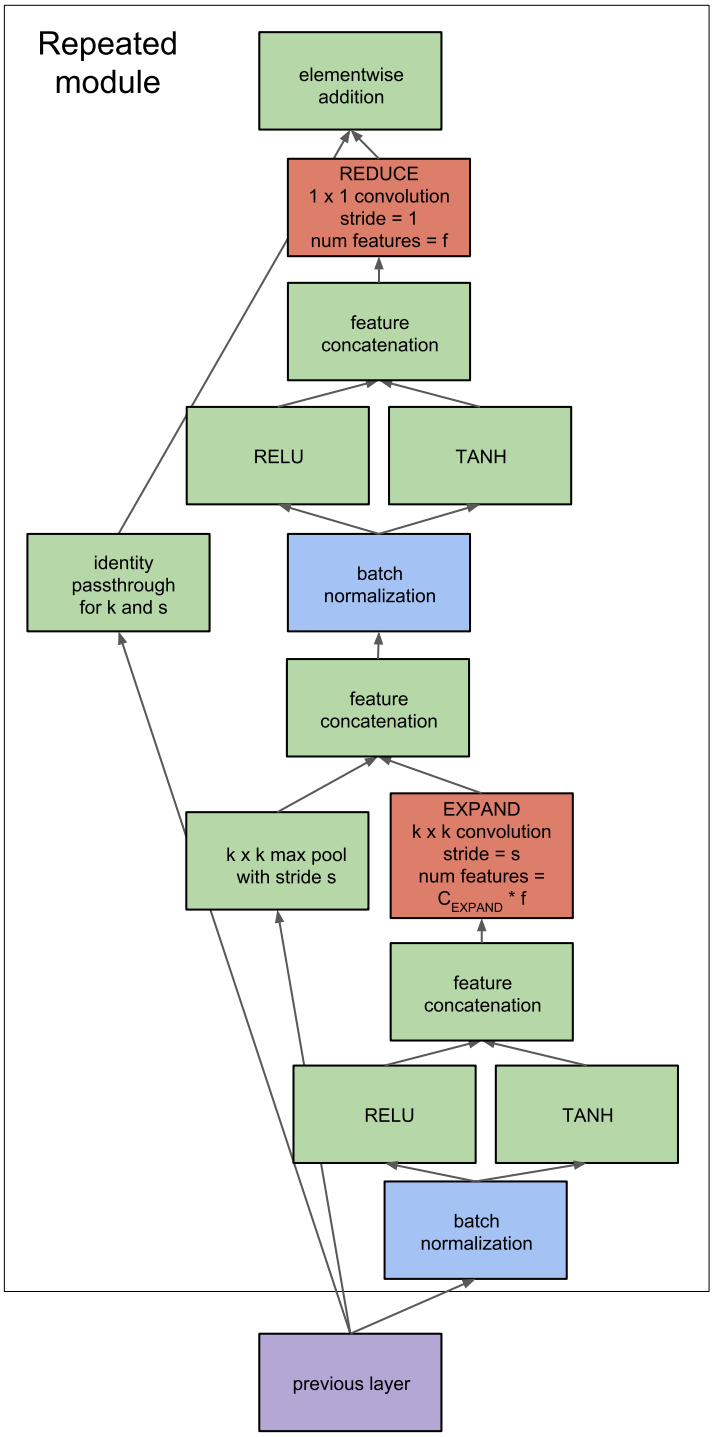
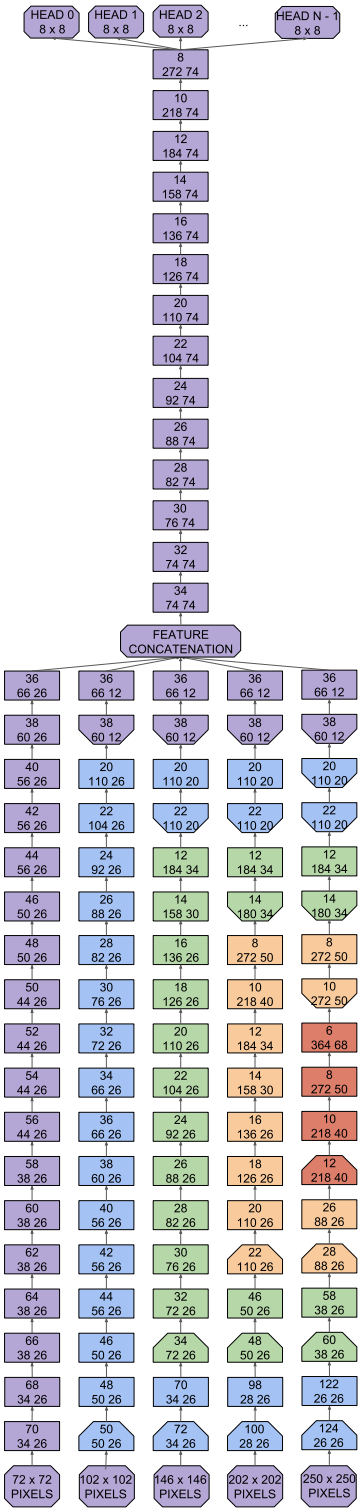
Methods S1 — Data and model details, related to Figures 1, 2, and 3



Methods S1 Figure 1: Generation of z-stacks of transmitted light images of unlabeled cells, related to Figures 1 and 3. To initially train the network and to test the predictions of the network, z-stacks of transmitted light images of a given microscope field were generated by collecting a total of 13 images: one approximately at the focal plane and an additional six images above and below that plane. In this example shown from Condition A, the 13 images in a stack were spaced 0.3 μm apart, spanning 3.6 μm along the z-axis. The location of each image relative to the central plane is given in microns by the numbers to left of the images. The outsets illustrate how different planes capture different information about the sample with some planes providing greater detail about intracellular structure and others providing more information about neurites and cell morphology. Scale bars are 40 μm. See also Figures 1 and 3.



Methods S1 Figure 2: The repeated module, the basic building block of the deep neural network, related to Figure 3. Data flows from the bottom to the top, along the indicated edges. Red operations contain variables to be learned, green operations have no trained variables, and blue operations are batch normalization [Ioffe and Szegedy, 2015]. This module is parameterized with three values: the number of features f, the size of the first convolution kernel k, and the stride s. C_{EXPAND} is a constant, which we set to 5.41 after hyperparameter tuning. It is used in one of three configurations: (1) in the in-scale configuration, $k = 3$ and $s = 1$; (2) in the down-scale configuration, $k = 4$ and $s = 2$; (3) in the up-scale configuration, $k = 4$, $s = 2$, the max pool is dropped, and the expand convolution is replaced with a transposed convolution [Zeiler et al., 2010], followed by a center crop to make the convolution transpose more space invariant. In this crop, activations within two rows or columns of the border are discarded. The convolutions and the max pooling are not zero-padded [Dumoulin and Visin, 2016], meaning they don't imputed missing activation values. See also Figure 3.



Methods S1 Figure 3: The deep neural network, the full statistical model used for label prediction, related to Figure 3. The rectangles and hexagons are the network modules: the rectangles are *in-scale*, the hexagons with flat bottoms are *down-scale*, and the hexagons with flat tops are *up-scale*. The octagons at the bottom are raw pixels read from the unlabeled image stack, and the octagons at the top are network heads, from which the predicted patches are derived for each fluorescent label. The colors correspond to the spatial scale of each particular module. Purple is the native scale, blue is 2× downscale, green is 4× downscale, orange is 8× downscale, and red is 16× downscale. The top number in each module is the number of rows and columns of its output layer. The bottom two numbers are the numbers of features in the module's expansion and reduction layers, respectively. The network reads from a concentric set of five square patches, ranging in size from 72×72 pixels to 250×250 pixels, processes each one independently, merges them, does more processing, then predicts a number of 8×8 patches. See also Figure 3.

References

[Dumoulin and Visin, 2016] Dumoulin, V. and Visin, F. (2016). A guide to convolution arithmetic for deep learning.

[Ioffe and Szegedy, 2015] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502. 03167*.

[Zeiler et al., 2010] Zeiler, M. D., Krishnan, D., Taylor, G. W., and Fergus, R. (2010). Deconvolutional networks. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2528–2535.